

Conditioning moments of singular measures for entropy maximization II: Numerical examples

Marko Budišić · Mihai Putinar

March 24, 2015

Abstract If moments of singular measures are passed as inputs to the entropy maximization procedure, the optimization algorithm might not terminate. The framework developed in [5] demonstrated how input moments of measures, on a broad range of domains, can be conditioned to ensure convergence of the entropy maximization. Here we numerically illustrate the developed framework on simplest possible examples: measures with one-dimensional, bounded supports. Three examples of measures are used to numerically compare approximations obtained through entropy maximization with and without the conditioning step.

MSC 30E05, 30E20, 41A46.

Keywords moments of singular measures, maximum entropy, Hilbert transform, moment problem, inverse problem

1 Introduction

Among all inverse problems, old and new, theoretical or applied, the reconstruction or, at least, approximation of a positive measure from moment data stands as a major topic of continued interest with history of a century and a half. Modern motivations for studying moment problems arise in continuum mechanics, statistics, image processing, control theory, geophysics, polynomial optimization, to name only a few. Added to these are purely theoretical aspects of convex algebraic geometry or harmonic analysis. While most cases deal either with measures absolutely continuous with respect to Lebesgue measure

To Ed Saff on the occasion of his seventieth birthday.

Department of Mathematics, University of Wisconsin, Madison, WI, USA E-mail: marko@math.wisc.edu · Department of Mathematics, University of Santa Barbara, Santa Barbara, CA, USA E-mail: mputinar@math.ucsb.edu, and School of Mathematics & Statistics Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom. E-mail: Mihai.Putinar@newcastle.ac.uk

or with purely atomic measures, the class of continuous singular measures has been poorly studied from the point of view of moment data inversion. The present note is a continuation of article [5] which developed a theoretical framework that includes all singular measures in the ubiquitous maximum entropy closure. Our aim here is to numerically demonstrate on several simple examples how input moment data can be conditioned using a nonlinear triangular transform, with the purpose of stimulating research into effects of such moment conditioning on the numerical accuracy, algorithm implementation, and even theoretical underpinnings of the maximum entropy method. For illustration, we choose the comfort of the measures with one-dimensional, bounded supports, although our previous work [5] additionally covered unbounded supports and multivariate moments. A forthcoming third article in this series will deal with inverse moment problems involving singular measures appearing in analysis of spectra of some concrete dynamical systems [4].

The amount of input moment data available is one of the distinguishing characteristics between theoretical and applied problems. Applied problems almost exclusively feature limited input data (*the truncated moment problem*) which means that the inversion is under-determined: there will be many measures whose moments match the inputs. The unique solution measure is obtained through *moment closure*, i.e., by posing an additional assumption on the input measure, which constrains the set of feasible measures in which the solution is searched for and ensures a unique solution. Requiring that the solution maximizes Boltzmann-Shannon entropy (MAXENT)¹ is an example of a closure that is particularly common in physical sciences and statistics [3].

Explanations of intuition behind MAXENT approach abound in literature, ranging from rigorous to philosophical [6, 9, 12, 17]. Informally, MAXENT probability measures are those that match the input moment data, but are otherwise unbiased towards any particular outcome. For our purposes, the important feature of MAXENT measures is that they are continuous, i.e., they have a density function. Their densities are in the form of exponential polynomials $e^{a_n x^n + a_{n-1} x^{n-1} + \dots}$ which means that the “rougher” the measure that generated the input data, the less likely it is for MAXENT to provide a good approximation from finitely-many moments. In the extreme, when input moments come from a singular measure, MAXENT approach may even fail to converge [5]. Singular measures, however, are hardly pathological if a broad range of domains is surveyed. In particular, they often appear as invariant measures of dynamical systems whose dynamics settles on a lower-dimensional attractor. In function theory they naturally arise in the study of boundary values of bounded analytic functions in a planar region [8].

In Section 2, we briefly re-state the basic theory from [5] needed to understand the single-variable, bounded-domain problem. Section 3 discusses three numerical problems implementation of the algorithm in MATLAB [18] and **Chebfun** [7]. Finally, we summarize the results in Section 4.

¹ Depending on the convention for choosing the sign of the entropy functional, both “minimization” and “maximization” appear in literature.

2 Analytical description

2.1 Three representations of a measure

Consider the bounded interval $[-\pi, \pi]$ and interpret it as the boundary $\partial\mathbb{D}$ of the open unit disk \mathbb{D} . Endow the interval with a measure μ with no assumptions on regularity of μ and, in particular, on existence of density $\mu'(\vartheta)$. The complex trigonometric moments of μ , also known as Fourier coefficients, are given by

$$\tau_\mu(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\vartheta} d\mu(\vartheta). \quad (1)$$

If the real-valued density $\mu'(\vartheta)$ is square-integrable, trigonometric moments are featured in its Fourier expansion:

$$\begin{aligned} \mu'(\vartheta) &= \sum_{n=-\infty}^{\infty} \tau_\mu(n) e^{in\vartheta} \\ &= \tau_\mu(0) + 2 \sum_{n=1}^{\infty} [\Re \tau_\mu(n) \cos(n\vartheta) - \Im \tau_\mu(n) \sin(n\vartheta)]. \end{aligned} \quad (2)$$

The measure μ induces the function $\dot{\mu}(z) : \mathbb{C} \rightarrow \mathbb{C}$ through the Cauchy integral²

$$\dot{\mu}(z) := \frac{i}{2\pi} \int_{-\pi}^{\pi} \frac{d\mu(\vartheta)}{1 - e^{-i\vartheta}z} = \frac{1}{2\pi} \int_{\zeta \in \partial\mathbb{D}} \frac{d\mu(\zeta)}{\zeta - z}. \quad (3)$$

The Cauchy transform is commonly defined on the real line as an integral of the Cauchy kernel $(x - z)^{-1}$ against a measure. As the expression (3) is equivalent to the classical Cauchy transform on the real line, we refer to it as the Cauchy transform as well. The function $\dot{\mu}(z)$ is known as the *analytic representation* of μ [14, §10.8], as $\dot{\mu}$ is analytic everywhere except on the unit circle $\partial\mathbb{D}$. Its derivatives at $z = 0$ are given by

$$\frac{d^k \dot{\mu}}{dz^k} = \frac{ik!}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-ik\vartheta} d\mu(\vartheta)}{(1 - e^{-i\vartheta}z)^{k+1}}, \quad \frac{d^k \dot{\mu}}{dz^k}(0) = ik! \tau_\mu(k), \quad (4)$$

so it follows that the power expansion of the analytic representation $\dot{\mu}(z)$ on the unit disk \mathbb{D} encodes the trigonometric moment sequence μ :

$$\dot{\mu}(z) = i \sum_{k=0}^{\infty} \tau_\mu(k) z^k. \quad (5)$$

When $\mu(\vartheta)$ has a Hölder-continuous density $\mu'(\vartheta)$, it is possible to recover it from $\dot{\mu}(z)$ by taking limits of $z \rightarrow e^{i\vartheta}$ along directions not tangential to the unit circle. Depending on the region approach with respect to the unit disk, we distinguish interior and exterior non-tangential limits, respectively denoted

² In literature, it is common to see $2\pi i$ in the denominator, which makes formulas in this paper slightly unorthodox.

by $\text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z)$ and $\text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z)$. For Hölder-continuous $\mu'(\vartheta)$, the non-tangential limits depend only on the region (interior/exterior) of the approach, but not on the precise direction. While taking i-lim is just a standard limit along the appropriate radius of \mathbb{D} , expressions (3) and (5) cannot be used directly to evaluate $\text{e-lim}_{z \rightarrow e^{i\vartheta}}$ where approach is through the complement of \mathbb{D} . However, it is possible to define the *associate function* $\dot{\mu}^\dagger(z) := \overline{\dot{\mu}(1/\bar{z})}$, analytic in \mathbb{D} , whose i-lim limit relates to the e-lim of $\dot{\mu}(z)$. Full details are available in [10, §14.2.I], but for our purposes the following formula for e-lim will suffice:

$$\text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z) = i\tau_\mu(0) + \overline{\text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z)}. \quad (6)$$

Pointwise values of limits are given by Plemelj–Sokhotski formulas³ [10, §14.2-IV]:

$$\begin{aligned} \text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z) &= \frac{i}{2}\mu'(\vartheta) + \frac{i}{2}\tau_\mu(0) - \frac{1}{2}\mathcal{H}\mu'(\vartheta) \\ \text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z) &= -\frac{i}{2}\mu'(\vartheta) + \frac{i}{2}\tau_\mu(0) - \frac{1}{2}\mathcal{H}\mu'(\vartheta) \end{aligned} \quad (7)$$

where $\mathcal{H}\mu'$ is the Hilbert transform of the density $\mu'(\vartheta)$

$$\mathcal{H}\mu'(\vartheta) = \frac{1}{2\pi} \oint_{-\pi}^{\pi} \cot \frac{\vartheta - \alpha}{2} \mu'(\alpha) d\alpha \quad (8)$$

evaluated using the principal value integral \oint . Subtracting the formulas we recover the density $\mu'(\vartheta)$ pointwise:

$$\begin{aligned} \mu'(\vartheta) &= i \left[\text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z) - \text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z) \right] \\ \mathcal{H}\mu'(\vartheta) &= - \left[\text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z) + \text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z) \right] + i\tau_\mu(0). \end{aligned} \quad (9)$$

Figure 1 illustrates how the concepts of measure μ , its moment sequence $\tau_\mu(k)$, and analytic representation $\dot{\mu}(z)$ relate. The stated relationships are well-documented in classical literature [8, 10, 14] and further references listed in [5].

As mentioned in Section 1, our goal is to specify a closure for the truncated moment problem: recover a representation of μ from a partial knowledge of moments τ_μ or, equivalently, truncation of power series of $\dot{\mu}(z)$. Two well-known examples of closures are the Padé approximation and the entropy maximization, both requiring additional assumptions on the (non-)existence of density $\mu'(\vartheta)$ and its smoothness. The Padé approximation can treat both singular and continuous measures μ , but at the cost that it always gives an atomic measure as the result which is, in practice, a poor approximation of any measure with a continuous density. On the other hand, the maximum entropy closure does

³ These formulas are commonly labeled by various subsets of names of Plemelj, Sokhotski, and Privalov, all of which have completed substantial work on them. Typically, these formulas are stated in a complex-argument form. The crucial step in conversion to angle form on the unit circle is the identity $(1 - e^{ix})^{-1} = [1 + i \cot(x/2)]/2$.

not converge when applied to moments generated by some singular measures, e.g., a point mass measure [5], thus not producing an approximation without an additional stopping criterion.

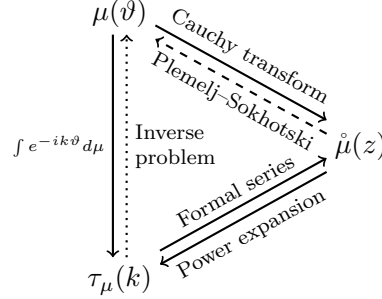


Fig. 1: Relations between the measure $\mu(\vartheta)$, its analytic representation $\hat{\mu}(z)$, and trigonometric moments $\tau_\mu(k)$. Plemelj-Sokhotski formulas (dashed) apply only when μ has a Hölder-continuous density function. A unique solution to the inverse problem (dotted) is available in certain cases when the full sequence $\tau_\mu(k)$, $\forall k \in \mathbb{Z}$ is known or, by closure, under additional restrictions on μ .

2.2 Even singular measures are represented by bounded densities

In this section we relate $\hat{\mu}$ to another analytic function with very similar properties but with an important improvement: it is represented by a Cauchy integral of a measure with a *bounded* density, even in cases of singular μ . We follow ideas of Markov [1] and Aronszajn and Donoghue [2] in this development.

For positive measures μ , $\hat{\mu}(z)$ maps \mathbb{D} into the upper half plane \mathbb{H}^+ , which is easy to demonstrate by computing the imaginary part of the integral kernel $i/(1 - e^{-i\vartheta}z)$. The consequence is that the principal branch of the angle⁴ $\text{Arg } \hat{\mu}$ is a real-valued *bounded and positive* function $\text{Arg } \hat{\mu}(z) : \mathbb{D} \rightarrow [0, \pi)$.

The class of complex-valued analytic functions of \mathbb{D} with a positive real part is known as the *Carathéodory class*. When $F(z)$ is Carathéodory and its real part is furthermore bounded by a positive constant F^+ , the Riesz-Herglotz theorem ([5], [10, §12.10]) asserts that $F(z)$ has an integral representation through measure $\phi(\vartheta)$ on the circle $\partial\mathbb{D}$:

$$F(z) = i\Im F(0) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\vartheta} + z}{e^{i\vartheta} - z} d\phi(\vartheta). \quad (10)$$

⁴ We use the following cartesian/polar notation for any $w \in \mathbb{C}$: $w = \Re w + i\Im w = |w| \exp(i \text{Arg } w)$, where the principal branch of argument is $\text{Arg} : \mathbb{C} \rightarrow [-\pi, \pi)$. Additionally, the complex conjugate is denoted by \bar{w} .

The integral in (10) is very closely related to the analytic representation (3) of the measure ϕ as it can be rewritten to give

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\vartheta} + z}{e^{i\vartheta} - z} d\phi(\vartheta) = -\tau_{\phi}(0) - 2i\check{\phi}(z). \quad (11)$$

We can, therefore, re-state the representation formula for $F(z)$ as

$$F(z) = i\Im F(0) - \tau_{\phi}(0) - 2i\check{\phi}(z). \quad (12)$$

Riesz–Herglotz theorem additionally establishes that the measure ϕ is continuous, i.e., there exists an integrable density function $\phi'(\vartheta)$. Furthermore, the bounds on the real part of $F(z)$, $\Re F(z) \in [0, F^+)$, also pointwise bound the density, implying $\phi'(\vartheta) \in [0, F^+)$.

As shown earlier, $\text{Arg } \check{\mu}(z) = -i \text{Log } \check{\mu}(z)$ is a positive real-valued function bounded by $[0, \pi)$, Riesz–Herglotz machinery can be applied to the principal branch of logarithm of $\mu(z)$. For any $M \geq 0$ (to be chosen later) set

$$F(z) := -i \text{Log} [\check{\mu}(z) + iM]. \quad (13)$$

At $z = 0$ we obtain

$$\begin{aligned} F(0) &= -i \text{Log} [\check{\mu}(0) + iM] = -i \text{Log} [i(\tau_{\mu}(0) + M)] \\ &= \pi/2 - i \log [\tau_{\mu}(0) + M], \end{aligned} \quad (14)$$

which implies that the Riesz–Herglotz representation is

$$F(z) = -i \log [\tau_{\mu}(0) + M] - \tau_{\phi}(0) - 2i\check{\phi}(z). \quad (15)$$

Equating (13) and (15) at $z = 0$, we find that the zeroth moment of ϕ is always $\tau_{\phi}(0) = \pi/2$, for any measure μ .

From equivalence of (13) and (15) we can solve for $\check{\mu}(z)$ to obtain its *analytic phase representation* in terms of $\check{\phi}(z)$:

$$\check{\mu}(z) = -iM - i[\tau_{\mu}(0) + M] \exp[2\check{\phi}(z)], \quad (16)$$

for any choice of $M \geq 0$.

The representation formula (16) establishes a connection between an arbitrary measure μ , which can be singular, and a measure ϕ with a bounded density function ϕ' . As $\check{\phi}$ arises as an argument of a complex exponential in (16), we refer to the measure ϕ and density ϕ' as the *phase measure* and *phase density*, respectively, of the measure μ .

Relations between three representations (Figure 1) apply to the phase measure ϕ as well, with the important distinction that the density ϕ' exists and is bounded, which brings us a step closer to applicability of Plemelj–Sokhotski formulas for reconstruction. While continuity of ϕ' cannot be theoretically expected, its assured boundedness makes it possible to approximate ϕ' by a continuous function, to which Plemelj–Sokhotski formulas then naturally apply.

Figure 2 illustrates the relations between discussed objects. In the next three sections, we operate in an applied context and describe how:

- (a) phase moments $\tau_\phi(k)$ are obtained from $\tau_\mu(k)$ when $k = 0, 1, \dots, K \leq \infty$,
- (b) $\phi'(\vartheta)$ is approximated by a continuous function from $\tau_\phi(k)$ via entropy maximization, and
- (c) μ is approximated using $\phi'(\mu)$ and its Hilbert transform $\mathcal{H}\phi'(\mu)$.

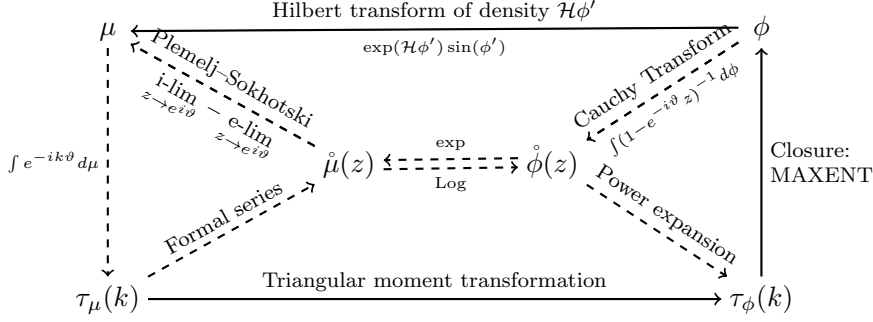


Fig. 2: Relations between measures μ and ϕ , their moments, and their analytic representations. Full arrows represent steps actually performed in our implementation, while dashed arrows represent analytical justifications that are not numerically evaluated (cf. Figure 1).

2.3 A triangular transformation conditions the moment sequence

As the expression (16) contains analytic functions, it is obvious that the full sequence of phase moments $\tau_\phi(k), \forall k \in \mathbb{Z}$ can be computed from the sequence $\tau_\mu(k), \forall k \in \mathbb{Z}$. A subtler point is that to compute the first K phase moments, $\tau_\phi(k), 0 \leq k \leq K-1$, we need just the first K moments $\tau_\mu(k), 0 \leq k \leq K-1$. The analytic phase representation (16) connects power expansions of $\dot{\mu}$ and $\dot{\phi}$ at $z = 0$

$$M + \sum_{n=0}^{\infty} \tau_\mu(n) z^n = -[\tau_\mu(0) + M] \exp \left[2i \sum_{k=0}^{\infty} \tau_\phi(k) z^k \right].$$

$$\text{Log} \left[M + \sum_{n=0}^{\infty} \tau_\mu(n) z^n \right] = \pm i\pi + \log[\tau_\mu(0) + M] + 2i \sum_{k=0}^{\infty} \tau_\phi(k) z^k.$$

The choice of the sign is made by evaluation at $z = 0$ with $\tau_\phi(0) = \pi/2$, resulting in

$$\sum_{k=0}^{\infty} \tau_\phi(k) z^k = \frac{\pi}{2} + \frac{i}{2} \log[\tau_\mu(0) + M] - \frac{i}{2} \text{Log} \left[M + \sum_{n=0}^{\infty} \tau_\mu(n) z^n \right]. \quad (17)$$

For the choice of $M = 1$, Log can be directly expanded into formal Mercator series $\text{Log}(1 + z) = -\sum_{n=1}^{\infty} \frac{(-1)^n}{n} z^n$:

$$\sum_{k=0}^{\infty} \tau_{\phi}(k) z^k = \frac{\pi}{2} + \frac{i}{2} \log[\tau_{\mu}(0) + 1] + \frac{i}{2} \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left[\sum_{n=0}^{\infty} \tau_{\mu}(n) z^n \right]^k. \quad (18)$$

Alternatively, we can extract $M + \tau_{\mu}(0)$ from the argument of Log in (17) to obtain:

$$\begin{aligned} \sum_{k=0}^{\infty} \tau_{\phi}(k) z^k &= \frac{\pi}{2} - \frac{i}{2} \text{Log} \left[1 + \sum_{n=1}^{\infty} \frac{\tau_{\mu}(n)}{M + \tau_{\mu}(0)} z^n \right] \\ &= \frac{\pi}{2} + \frac{i}{2} \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left[\sum_{n=1}^{\infty} \frac{\tau_{\mu}(n)}{M + \tau_{\mu}(0)} z^n \right]^k \end{aligned} \quad (19)$$

While expressions (18) and (19) are largely similar, we draw the attention to the lower bound of the inner sum \sum_n , which runs from $n = 0$ in (18) and from $n = 1$ in (19). In the remainder of the paper, we will discuss only $M = 0$ case of formulas, corresponding to (19).

In either (18) or (19) the exponents of z are non-negative, so coefficient $\tau_{\phi}(k)$ cannot depend on any coefficients $\tau_{\mu}(n)$ in which $n > k$. It follows that to compute a finite number of $\tau_{\phi}(k)$, only a finite number of $\tau_{\mu}(n)$ is needed. Transformations of sequences $\mathbf{u} \mapsto \mathbf{v}$ in which element v_n depends only on elements u_1, u_2, \dots, u_n are referred to as “triangular” in analogy to linear transformations of this kind, which are represented by lower-triangular matrices

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} * & & \\ * & * & \\ \vdots & \ddots & \ddots \\ * & * & \dots & * \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}.$$

Triangular transformations of sequences are important in calculations where only a finite number of input elements are available, as it implies that at least the initial part of the output sequence can be obtained exactly.

2.4 A continuous approximation of the phase density using entropy maximization

Assume that a finite set of complex numbers $\tau(k)$, $k = 0, \dots, K - 1$ is chosen such that there exists at least one measure $\rho'(\vartheta) d\vartheta$ whose truncated moment sequence matches numbers $\tau(k)$:

$$\tau_{\rho}(k) = \tau(k), \quad k = 0, \dots, K - 1. \quad (20)$$

The set of all measures that satisfy the truncated moment condition (20) is called the *feasible set* \mathcal{F} . Whenever $K < \infty$, the feasible set is uncountable and the inverse problem, i.e., selecting a single measure that corresponds to the truncated moment conditions, does not have a unique solution.

Entropy maximizations (MAXENT) are a broad class of closures which are well-understood for various domains of measures [3, 13]. We discuss the simplest case of Boltzmann-Shannon entropy maximization on the continuous domain: real-valued measures on a real, bounded interval. Since the measure is real-valued, truncated moment conditions (20) additionally define the negative-order moments of orders up to $-(K-1)$ through relation $\tau_{-k} = \bar{\tau}_k$. The entropy maximization solves the inverse problem by selecting the measure that lies in the feasible set and maximizes the Boltzmann-Shannon entropy functional:

$$\arg \max_{\rho \in \mathcal{F}} \int \log \rho'(\vartheta) d\rho(\vartheta). \quad (21)$$

In the setup as simple as ours, the solution is easily found by the Lagrange multipliers technique, searching for the stationary point of the argument of the integral in the functional

$$\int \left[\rho'(\vartheta) \log \rho'(\vartheta) + \frac{1}{2\pi} \sum_{k=-K}^K \alpha_k \rho'(\vartheta) e^{ik\vartheta} \right] d\vartheta, \quad (22)$$

where α_k are the Lagrange multipliers. As the result, we obtain the density

$$\rho'(\vartheta) = \exp \left(\sum_{k=-K}^K \alpha_k e^{ik\vartheta} \right), \quad \alpha_{-k} = \bar{\alpha}_k \quad (23)$$

for the solution of the inverse problem (with some constants absorbed into α_k). The procedure can be made quite general, as in [3, 13]; here, we work with (23) that states that solutions with maximal entropy are always exponential trigonometric polynomials whose degree matches the number of truncated moment conditions. Practically, this implies that the entropy maximization can be stated as a dual nonlinear optimization problem where coefficients α_k are varied until the moments of measure ρ match the moment conditions. Applying entropy maximization defines higher moments $\tau_\rho(k)$, $|k| > K$ through a recursive equation that depends on the coefficients α_k as derived in [5].

The entropy maximization is attractive for several reasons. First, in the limit when a full moment sequence is known, MAXENT recovers the unique measure that generates the moment sequence [13]. From the practical standpoint, resolution of a truncated moment problem is attractive in physical sciences since *it always produces a smooth and positive density* through little more than a convex optimization process. Unfortunately, not all finite sets of complex numbers τ_k result in truncated moment conditions that can be matched by the MAXENT ansatz, as is easy to show by studying an atomic probability measure $\delta(\vartheta)$ [5].

2.5 The density is computed from the phase using the Hilbert transform

Using the entropy maximization as the closure of the truncated moment problem for ϕ allows us to assume that we have a smooth density $\phi'(\vartheta)$ featured

in the analytic phase representation (16)

$$\dot{\mu}(z) = -iM - i[\tau_\mu(0) + M] \exp \left[\frac{i}{\pi} \int_{-\pi}^{\pi} \frac{\phi'(\vartheta) d\vartheta}{1 - e^{-i\vartheta} z} \right].$$

Plemelj–Sokhotski formulas (7) can be used to reconstruct the continuous density of μ as a difference of non-tangential limits of $\dot{\mu}(z)$. While our original measure μ did not necessarily have a continuous density, the use of entropy maximization to obtain a smooth phase density has the effect of regularizing μ as well.

Due to analyticity of the exponential function, non-tangential limits of $\dot{\mu}(z)$ can be stated in terms of limits of $\dot{\phi}(z)$

$$\begin{aligned} \text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z) &= -iM - i[\tau_\mu(0) + M] \exp \left[2 \text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\phi}(z) \right], \\ \text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z) &= i\tau_\mu(0) + \overline{\text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\mu}(z)} \\ &= i\tau_\mu(0) + iM + i[\tau_\mu(0) + M] \exp \left[2 \text{i-lim}_{z \rightarrow e^{i\vartheta}} \overline{\dot{\phi}(z)} \right] \\ &= i\tau_\mu(0) + iM - i[\tau_\mu(0) + M] \exp \left[2 \text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\phi}(z) \right], \end{aligned} \quad (24)$$

where we used (6) and $\tau_\phi(0) = \pi/2$ to relate the interior and exterior non-tangential limits.

Taking the difference of non-tangential limits for $\dot{\mu}(z)$ and factoring the exponential yields *the inversion formula*

$$\mu'(\vartheta) = -2M - \tau_\mu(0) + 2[\tau_\mu(0) + M] e^{-\mathcal{H}\phi'(\vartheta)} \sin \phi'(\vartheta). \quad (25)$$

Note that the choice of $M \geq 0$ implicitly affects moments $\tau_\phi(k)$, through power expansion relation (17), and, consequently, the recovered $\phi'(\vartheta)$.

While the inversion formula (25) seemingly involves another singular integral through the Hilbert transform, this transformation has a simple spectral representation. If a general function f is expanded into Fourier series as $f(\vartheta) = \sum_{n=-\infty}^{\infty} \tau_n e^{in\vartheta}$, then the expansion of its Hilbert transform $\mathcal{H}f$ is

$$\mathcal{H}f(\vartheta) = -i \sum_{n=-\infty}^{\infty} \text{sgn}(n) \tau_n e^{in\vartheta}. \quad (26)$$

In other words, the Hilbert transform can be efficiently computed using commonly available Fast Fourier Transform (FFT) and inverse FFT algorithms.⁵

⁵ Alternatively, if Chebyshev expansion of $f(\vartheta)$ is given $f(\vartheta) = \sum_n c_n T_n(\vartheta)$, then the Hilbert transform of a rescaled function $f(\vartheta)(1 - \vartheta^2)^{-1/2}$ has an expansion $\mathcal{H}[f(\vartheta)(1 - \vartheta^2)^{-1/2}] = -\sum_n c_n U_{n-1}(\vartheta)$, where T_n and U_n are Chebyshev polynomials of the first and second kind, respectively. Additional formulas for an explicit expansion of $\mathcal{H}f$ into Chebyshev polynomials of the first kind are derived over several different domains in [16].

2.6 An essential example: a point mass measure

Although the high-level concepts of the moment conditioning are easily understood from Figure 2, the detailed calculations require delicate choices of branches of inverse functions, which can be performed explicitly on an atomic measure μ . When $\mu(\vartheta)$ is a point-mass at $\vartheta = a$, i.e., $d\mu(\vartheta) = \delta_a(\vartheta)d\vartheta$, most of the expressions can be determined analytically as integrals against the Dirac δ_a , which simplifies to evaluations of the integrand at $\vartheta = a$.

Trigonometric moments are given by $\tau_\mu(k) = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-ik\vartheta} \delta_a(\vartheta) d\vartheta = e^{-ika}/(2\pi)$ and the analytic representation $\dot{\mu}(z)$ on \mathbb{D} is

$$\dot{\mu}(z) = \frac{i}{2\pi} \int_{-\pi}^{\pi} \frac{\delta_a(\vartheta) d\vartheta}{1 - e^{-i\vartheta} z} = \frac{i}{2\pi} \frac{1}{1 - e^{-ia} z} = i \sum_{n=0}^{\infty} \frac{e^{-ian}}{2\pi} z^n. \quad (27)$$

To calculate $\dot{\phi}(z)$, use the analytic phase representation (16) and write

$$\begin{aligned} \exp[2\dot{\phi}(z)] &= \frac{i\dot{\mu}(z) - M}{\tau_\mu(0) + M} = \frac{2\pi i}{1 + 2\pi M} \dot{\mu}(z) - \frac{2\pi M}{1 + 2\pi M} \\ &= \frac{1}{1 + 2\pi M} \cdot \frac{1 + 2\pi M - 2\pi M e^{-ia} z}{e^{-ia} z - 1} \end{aligned}$$

At this point, choose $M = 0$ which reduces the representation to

$$\begin{aligned} \exp[2\dot{\phi}(z)] &= \frac{1}{e^{-ia} z - 1}. \\ \dot{\phi}(z) &= -\frac{1}{2} \text{Log}[e^{-ia} z - 1]. \end{aligned}$$

where the branch $\text{Log}[-1] = -i\pi$ is preferred so that $\dot{\phi}(z)$ would match the value of the zeroth moment $\tau_\phi(0) = \pi/2$ as calculated in (15).

As $\dot{\phi}(z)$ is analytic inside \mathbb{D} , calculating $\text{i-lim}_{z \rightarrow e^{i\vartheta}}$ is done trivially by substituting the limit point. Taking e-lim requires using the associate function, as in (6), to obtain

$$\begin{aligned} \text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\phi}(z) &= -\frac{1}{2} \text{Log}[e^{i(\vartheta-a)} - 1] \\ \text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\phi}(z) &= -\frac{1}{2} \text{Log}[e^{-i(\vartheta-a)} - 1] + i\frac{\pi}{2}. \end{aligned} \quad (28)$$

By Plemelj–Sokhotski (9), we obtain

$$\begin{aligned} \phi'(\vartheta) &= -i \left[\text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\phi}(z) - \text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\phi}(z) \right] = \frac{i}{2} \text{Log} \frac{e^{i(\vartheta-a)} - 1}{e^{-i(\vartheta-a)} - 1} - \frac{\pi}{2} \\ &= -\text{Arccot} \cot \frac{\vartheta - a}{2} + \pi = -\text{mod} \left(\frac{\vartheta - a}{2}, \pi \right) + \pi. \\ \mathcal{H}\phi'(\vartheta) &= - \left[\text{i-lim}_{z \rightarrow e^{i\vartheta}} \dot{\phi}(z) + \text{e-lim}_{z \rightarrow e^{i\vartheta}} \dot{\phi}(z) \right] + i\tau_\phi(0) \\ &= \frac{1}{2} \text{Log}[2 - e^{i(\vartheta-a)} - e^{-i(\vartheta-a)}] = \text{Log} \left[2 \left| \sin \frac{\vartheta - a}{2} \right| \right]. \end{aligned} \quad (29)$$

In all calculations above, care must be taken to select appropriate branches that ensure that $\phi'(\vartheta) \in [0, \pi]$, as stipulated by the Riesz–Herglotz formula.

To evaluate $\mu'(\vartheta)$, we use the inversion formula (25)

$$\begin{aligned}\mu'(\vartheta) &= -\tau_\mu(0) + 2\tau_\mu(0)e^{-\mathcal{H}\phi'(\vartheta)} \sin \phi'(\vartheta) \\ &= -\tau_\mu(0) + \tau_\mu(0) \frac{1}{\left|\sin \frac{\vartheta-a}{2}\right|} \sin \left[-\operatorname{mod} \left(\frac{\vartheta-a}{2}, \pi \right) + \pi \right] \\ &= -\tau_\mu(0) + \tau_\mu(0) \frac{\sin \operatorname{mod} \left(\frac{\vartheta-a}{2}, \pi \right)}{\left|\sin \frac{\vartheta-a}{2}\right|}.\end{aligned}$$

Note that for all $\vartheta \neq 0$ it holds that $\sin \operatorname{mod} \left(\frac{\vartheta-a}{2}, \pi \right) = \left|\sin \frac{\vartheta-a}{2}\right|$, so $\mu'(\vartheta \neq 0) \equiv 0$. However, since $\int d\mu(\vartheta) = 1$, it has to hold that $\mu'(\vartheta) = \delta_a(\vartheta)$.

3 Numerical examples

To demonstrate the effect that the conditioning has on solution of the inverse moment problem, we implemented the entire procedure as a MATLAB code, relying heavily on **Chebfun** [7, 19] package. **Chebfun** represents functions constructively using their Chebyshev or Fourier coefficients, instead of storing their pointwise values. Consequently, operations involving integrals and integral transforms are quicker and more precise than if it was implemented naively.

To condition the moments, we computed powers of the truncated analytic representation (17) directly, through discrete convolution of polynomial coefficients. Alternatively, it is possible to specify coefficients recursively using the Miller–Nakos algorithm [15], which would become preferable for implementing a multivariate generalization of this method [5].

The solution to the truncated moment problem using entropy maximization is cast as an unconstrained convex optimization problem and solved using MATLAB's internal derivative-free quasi-Newton optimization method (**fminunc**) [18]. The optimization function is the square-error in first K moments $\mathcal{E} = \sum_{k=0}^{K-1} |\tau(k) - \tau_\rho(k)|^2$, where $\tau(k)$ represent input data, and τ_ρ moments of the density in the maximum entropy ansatz (23). The optimization parameters are real and imaginary parts of complex coefficients α_k for $0 \leq k < K-1$, where coefficients $k < 0$ are obtained by imposing $\overline{\alpha_k} = \alpha_{-k}$, as in (2).

Since **Chebfun** represents functions internally using their Fourier coefficients, implementing the Hilbert transform reduces to a multiplication by a vector of constants. To transform the coefficients, the relation (26) is applied directly while **Chebfun** takes care of pointwise evaluation.

Numerical algorithm is applied to three different measures μ :

(a) a point mass measure,

$$\mu'(\vartheta) = \delta_a(\vartheta), \quad (30)$$

(b) a continuous measure with smooth (superposition of Gaussians) density

$$\mu'(\vartheta) = 5e^{-(5\vartheta-10)^2} + e^{-(5\vartheta+7.5)^2}, \quad (31)$$

(c) a continuous measure with discontinuous (rectangular) density

$$\mu'(\vartheta) = \begin{cases} 1 & x \in \left(-\frac{\pi-1}{2}, \frac{\pi+1}{2}\right) \\ 0 & \text{otherwise} \end{cases}. \quad (32)$$

All the measures were further normalized to the unit mass.

Assume that only first $K = 20$ initial moments $\tau_\mu(k)$ of the “true” density μ are available. Two approximations, μ'_U and μ'_C , to the initial density are obtained using, respectively, unconditioned MAXENT procedure and the conditioned MAXENT procedure (full arrows in Figure 2). In the text, we will refer to them as U-MAXENT and C-MAXENT. Settings of the optimization algorithm that constructs maximum entropy density are kept the same in both procedure.

While theoretically U-MAXENT may not converge for singular measures, numerical implementations have a number of stopping criteria which prevent the code from running forever. In our case, MATLAB’s stock optimization code terminated by reporting that the local minimum was numerically found in all but one example: when applying U-MAXENT to the singular measure, it stopped due to inability to reduce the moment error, without confirming numerically that the local minimum was found.

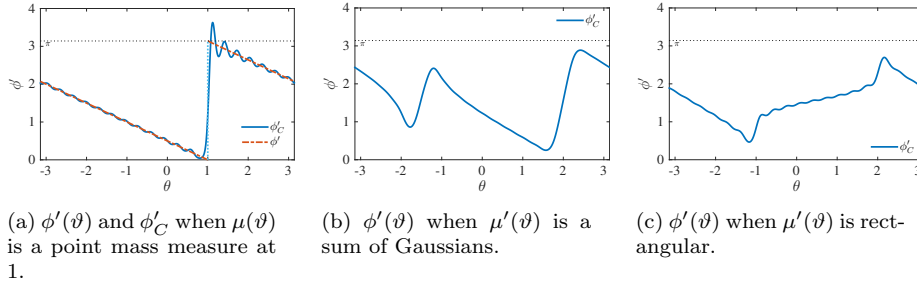


Fig. 3: Reconstructed phase densities ϕ'_C for three studied “ground” measures μ .

Before we discuss differences between U- and C-MAXENT, Figure 3 shows what reconstructed phase densities ϕ' look like for all three studied measures. While the phase density of δ_a can be easily computed analytically (Section 2.6), we have not attempted to do so for the remaining two densities. Loosely speaking, $\phi'(\vartheta)$ appear to resemble “integrals” of $\mu'(\vartheta)$ as singularities correspond to finite discontinuities, discontinuous changes are smoothed out, and constant regions are transformed into slopes. Intuitively, this should make ϕ' smoother

than μ' and therefore more amenable to reconstruction by entropy maximization.

Since the reconstruction algorithm was performed in the Fourier space, it is not surprising that the Gibbs phenomenon [11] at discontinuities results in overshooting of upper bound at π , with effects on μ_C discussed later in this section. Note that the maximum entropy ansatz $e^{P(\vartheta)}$ produces overshooting only in the positive direction, because Fourier coefficients of exponent $P(x)$ are the optimization parameters, and not the full sequence of Fourier coefficients of the density $e^{P(\vartheta)}$.

Figure 4 compares moments of reconstructed densities, μ'_C and μ'_U , with moments of μ' . Moment errors of μ'_C at orders $k < K = 20$ are lower by at least one order of magnitude than the same errors for μ'_U in all studied examples. At orders $k > K$, however, the advantage of μ'_C is less clear. Moment errors of μ'_U remain roughly constant when k is increased beyond K , except in the case of the discontinuous “box” density, where the error sharply increases. On the other hand, μ'_C moment errors experience similar sharp increase in all three studied examples. Nevertheless, after the increase, the errors in moments of μ'_C and μ'_U are mostly comparable. The exception is the case of a point mass measure, where U-MAXENT outperforms C-MAXENT for $k > K$.

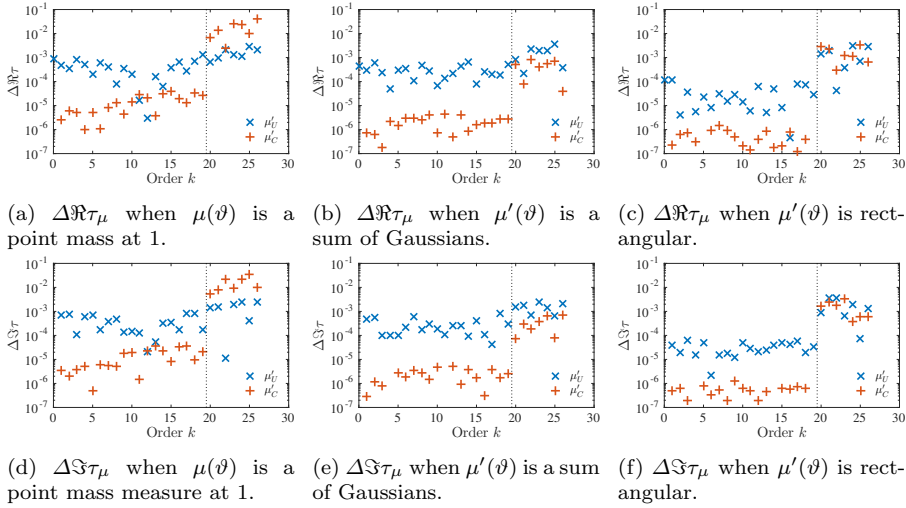


Fig. 4: Errors of moments of measures reconstructed using MAXENT with and without conditioning, resp., μ_C and μ_U , with respect to moments of the “true” measure μ for the point mass measure (30), a sum-of-Gaussians density (31), and a discontinuous density (32).

From pointwise comparison of densities in Figure 5, it can be seen that when μ' is smooth, like in the sum of Gaussians case, C-MAXENT provided a slightly better approximation than U-MAXENT. It is not clear, though,

whether that difference is practically relevant. On the other hand, the advantage to using U-MAXENT, when the process converges, is that its density remains positive, regardless of errors introduced. Reconstruction of densities in the case of δ_a and rectangular densities show that μ'_C crosses 0 into negative values. Both of these crossings are due to the Gibbs phenomenon, which is manifested on two different levels. The undershoot of the point mass reconstruction is a consequence of the Gibbs phenomenon in the reconstruction of the phase ϕ' , as can be seen in Figure 3, which propagates through $\sin \phi'(\vartheta)$ term in the inversion formula (25) and introduces negative values into reconstruction μ'_C . On the other hand, the undershoot in the reconstruction of the rectangular density does not have a counterpart in the phase density, but is likely a direct Gibbs effect, as the inversion formula (25) was implicitly evaluated using finitely many Fourier coefficients (see Table 1).

Example	μ'_U	μ'_C	ϕ'_C
Point mass	1709	747	343
Sum of Gaussians	291	253	193
Rectangular	389	343	237

Table 1: Number of Fourier coefficients internally chosen by **Chebfun** toolbox to represent densities.

4 Conclusions

Closures of inverse moment problems often require that the input moments correspond to an absolutely continuous measure, as is the case of the maximum entropy closure. If the input measure is singular, optimization algorithm maximizing the entropy may not terminate. In this case, the finite moment sequence of the measure can be conditioned so that the entropy maximization terminates and resolves the inverse problem [5]. This paper demonstrates that it is possible to numerically implement the conditioning procedure for the case of measures on a compact interval and that, indeed, conditioning of moments successfully allows the entropy maximization procedure to terminate.

Numerical results show that conditioning moments before passing them to entropy maximization consistently results in smaller errors between input moments and the moments of the reconstruction, as compared to the unconditioned entropy maximization. Comparing errors that an approximation makes *beyond* the first K moments shows less-definitive results: conditioned and unconditioned procedures result in similar level of errors in most cases, except when the measure is singular. In that case the unconditioned algorithm surprisingly performed better than the conditioned one. Pointwise comparison reveals that the final step (25) in the conditioned algorithm, inverting the effect of moment conditioning, exposes the procedure to numerical errors. In our

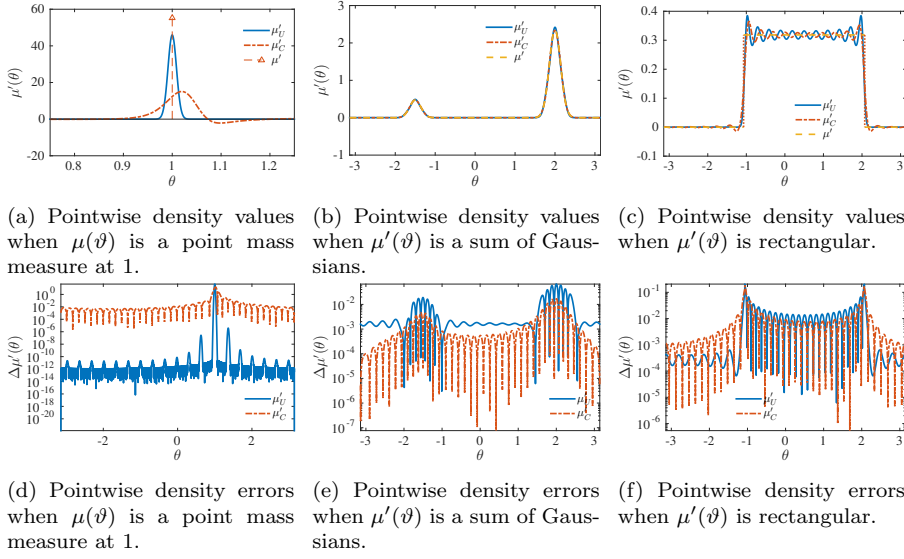


Fig. 5: Pointwise comparison of densities reconstructed using MAXENT with and without conditioning, resp., μ'_C and μ'_U , with respect to the densities of “true” measures μ for the point mass measure at 1 (30), a sum-of-Gaussians density (31), and a discontinuous density (32). δ_a is indicated by an arrow, and for it, the error was not computed at $\vartheta = a$ as it is always infinite.

implementation, this manifested as pointwise deterioration of the final approximation, due to the Gibbs phenomenon. The measures chosen for the examples were finitely determined by their moments, so there is no surprise that even the unconditioned entropy maximization performed well. In summary, this paper successfully demonstrated that implementing the conditioning procedure is practical, even though the used proof-of-concept algorithm may not be the best possible implementation.

Future numerical analysis of our modification to the entropy maximization would study the effects of error propagation through three major steps (Figure 2), as well as possibilities for mitigating those errors. Additionally, extensions of the procedure into higher-dimensional domains have been studied theoretically in [5], but the practical implementation would be an important achievement. Extending conditioning of moments is almost immediate, however the numerical implementation of the inversion step is less straightforward; additional work is needed to see how this obstacle can be overcome.

References

- [1] Naum I. Akhiezer and Mark G. Krein, *Some questions in the theory of moments*, translated by W. Fleming and D. Prill. Translations of Mathematical Monographs, Vol. 2, American Mathematical Society, Providence, R.I., 1962.

- [2] Nachman Aronszajn and William F. Donoghue Jr., *On exponential representations of analytic functions in the upper half-plane with positive imaginary part*, Journal d'Analyse Mathématique **5** (1956), no. 1, 321–388.
- [3] Jonathan M Borwein, *Maximum entropy and feasibility methods for convex and non-convex inverse problems*, Optimization: A Journal of Mathematical Programming and Operations Research **61** (2012), no. 1, 1–33.
- [4] Marko Budišić, Ryan M. Mohr, and Igor Mezić, *Applied Koopmanism*, Chaos: An Interdisciplinary Journal of Nonlinear Science **22** (December 2012), no. 4, 047510–1–33.
- [5] Marko Budišić and Mihai Putinar, *Conditioning moments of singular measures for entropy optimization I*, Indagationes Mathematicae **23** (December 2012), no. 4, 848–883.
- [6] Imre Csiszár, *Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems*, The Annals of Statistics **19** (1991), no. 4, 2032–2066.
- [7] Tobin A. Driscoll, Nicholas Hale, and Lloyd N. Trefethen, *Chebfun guide v.5.0.1*, Pafnuty Publications, 2014.
- [8] Gennadii M. Goluzin, *Geometric theory of functions of a complex variable*, Translations of Mathematical Monographs, Vol. 26, American Mathematical Society, Providence, R.I., 1969. MR0247039
- [9] Marian Jr. Grendár and Marian Grendár, *Maximum entropy: Clearing up mysteries*, Entropy **3** (2001), no. 2, 58–63.
- [10] Peter Henrici, *Applied and computational complex analysis*, Wiley Classics Library, vol. 1-3, Wiley-Interscience, New York, 1977.
- [11] Edwin Hewitt and Robert E. Hewitt, *The Gibbs-Wilbraham phenomenon: an episode in Fourier analysis*, Archive for history of Exact Sciences **21** (1979), no. 2, 129–160.
- [12] Edwin T Jaynes, *On the rationale of maximum-entropy methods*, Proceedings of the IEEE **70** (1982), no. 9, 939–952.
- [13] Michael Junk, *Maximum entropy for reduced moment problems*, Mathematical Models & Methods in Applied Sciences **10** (2000), no. 7, 1001–1025.
- [14] Frederick W. King, *Hilbert transforms*, Encyclopedia of Mathematics and its Applications, vol. 124, Cambridge University Press, Cambridge, 2009.
- [15] George Nakos, *Expansions of powers of multivariate formal power series*, The Mathematica Journal **3** (September 1993), no. 1, 45–47.
- [16] Sheehan Olver, *Computing the Hilbert transform and its inverse*, Mathematics of Computation **80** (2011), no. 275, 1745–1767.
- [17] Teddy Seidenfeld, *Entropy and uncertainty*, Philosophy of Science (1986).
- [18] The Mathworks, Inc, *MATLAB and optimization toolbox*, The MathWorks, Inc., Natick, Massachusetts, United States, 2014.
- [19] Lloyd N. Trefethen, *Approximation theory and approximation practice*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013. MR3012510